HW 5

Tara Hinton

12/29/2023

This homework is meant to give you practice in creating and defending a position with both statistical and philosophical evidence. We have now extensively talked about the COMPAS [1] data set, the flaws in applying it but also its potential upside if its shortcomings can be overlooked. We have also spent time in class verbally assessing positions both for an against applying this data set in real life. In no more than two pages [2] take the persona of a statistical consultant advising a judge as to whether they should include the results of the COMPAS algorithm in their decision making process for granting parole. First clearly articulate your position (whether the algorithm should be used or not) and then defend said position using both statistical and philosophical evidence. Your paper will be grade both on the merits of its persuasive appeal but also the applicability of the statistical and philosohpical evidence cited.

It is important to understand the moral implications and statistical grounds of the COMPAS algorithm before using it. In this analysis, I argue that the use of COMPAS undermines critical moral issues of fairness and equality.

Distributive Justice

We can employ philosophical frameworks of distributive justice and fairness to understand the implications of COMPAS. In this analysis, I explore how two frameworks of justice – justice as merit and John Rawls's Difference Principle – indicate moral issues with COMPAS use.

One approach to distributive justice treats fairness principally as a function of merit, where the allocation of resources depends on constructions of merit (progress, understanding, hard work, etc.). In the case of parole hearings, the use of COMPAS fails to meet our understandings of equality as merit. Statistically, our measure for fairness as merit is expressed in terms of equalized-odds, which indicate how often, when, and in what ways COMPAS fails to predict recidivism accurately. Using equalized-odds, we find differences in true positive and false positive rates across white and Black individuals. This is unfair because algorithm predictions produce more false positives among Black inmates than white inmates. In application, the algorithm alone does not consider the hard work, improvement, or propensity of the individual it examines. COMPAS invisibilizes evidence of rehabilitation, one factor that is critical to parole decisions in courtrooms today. This merit factor, here understood as evidence of managing harmful behavior include expert evaluations of an inmate's progress and fundamentally human observations. In effect, those applying for parole always have and always will receive the same prediction of recidivism from COMPAS, regardless of the ways in which they spent their time in jail. A virtue ethics framework, which acknowledges the ability of humans to grow and transform, would find this algorithm in fundamental opposition to what we would hope to cultivate in society. Its use, even as supporting evidence, deprives incarcerated individuals of the human prerogative to grow and change from the genesis of the algorithm at the variable level. From here, when we acknowledge that the fairness of individual parole necessarily involves a component of merit, we can begin to understand broader, social dimensions of equality.

John Rawls's Difference Principle posits that when vulnerabilities arise as a result of differences, resource allocation should prioritize the least advantaged. Here, need rather than merit determines fairness. Currently, there is no specific statistical metric of justice as need. Still, with its disparate impact on Black inmates,

---

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[2] knit to a pdf to ensure page count

COMPAS does not benefit the least advantaged or protect the least vulnerable. Proponents of COMPAS might argue that systemic human bias against minority populations already exists throughout the system – what makes an algorithm worse? Under Rawls's Difference Principle, we would reject the use of anything that would cause further harm to minority populations. Since we know with certitude that COMPAS fails statistical parity, equalized-odds, and disparate impact, it would be highly unethical to use it, while the presence of bias in a judge is not certain. Further, its use of training data impacted by historical policing, brutality, and redlining risks "twisting the knife" into minority populations by reproducing historical patterns of harm.

Conclusion: Kant and Scaling Up Recidivism Algorithms

Despite pervading moral issues, recidivism algorithms have made their way into justice system. Kant's universal maxim provides a unique way to view this scaled up use of COMPAS and similar algorithms. We can consider the impact of universally applying COMPAS use to all court systems and parole hearings. In such a world where COMPAS is a staple in decision-making, perhaps individual, contextual parole hearings become irrelevant. A court's judgement would be influenced then, in all cases, but an only moderately accurate model (61% approx.) with disparate impacts. This bias would produce a self-defeating "justice" system that would violate Kant's universal maxim. Statistically and morally, need-based and merit-based understandings of justice cannot uphold the use of COMPAS in a parole hearing. Should we wish for a more just justice system, we have to make better attempts for account for the nuance of human transformation and agency.

*what is the philosophical framework 1. Reason 1: Bias is just not morally right or equitable; we can argue that human judgements are bias too, but this doesn't justify the inclusion of a similarly incompetent system 2. Reason 2: 3. Reason 3: COMPAS implies that a fixed set of historical outcomes determined by a stable set of variables accounts for behavior. In effect, those applying for parole always have and always will receive the same predicition of recidivism from COMPAS, regardless of the ways in which they spent their time in jail. A virtue ethics framework, which acknowleges the ability of humans to grow and transform, would find this algorthm in fundamental opposition to what we would hope to cultivate in society. Its use, even as supporting evidence, deprives incarcerated individuals of the human perogative to grow and change from the genesis of the algorithm at the variable level. Who says historical recidivism rates are the proper way to judge "deservedness" of parole? When we examine the statistic and resulting moral error of COMPAS, we are confronted by a pixelated image of the defendant

Conclusion:

Our justice system should seek to build safer, healthier communities. COMPAS, while well-intentioned, cannot account for the nuance of human transformation and agency.